KUL 2020 - Presentation Sheila Galt, CLS, Chalmers

A coarse look at course evaluations - coupling to course quality and gender bias

Short summary:

What do student course evaluations really tell us about course (and teacher) quality? Let's look at possible problems (example: gender bias) and other evaluation methods (example: teaching practices inventory). Data on "overall impression", suggests that Chalmers students favour courses given by male examiners when it comes to "top notch" ratings.

Keywords:

student evaluations of teaching (SETs), teaching practice inventories (TPIs), gender bias, active learning, lecture-based courses

Introduction

In striving to move closer to world-class education, it can be useful to take a moment to investigate the concept of course quality. What do we mean, how do we work towards higher quality, how do we know if what we do really works, and how do we effectively measure course quality? Can we separate course quality from teacher quality? Might there even be gender bias effects in our present methods of measuring course quality? Many questions need discussion, but we will focus on two issues as described below.

In this paper we first discuss a number of possible problems with the use of student evaluations of teaching (SETs) as a main indicator of course (and sometimes teacher) quality. We then look into one such possible problem, gender bias, in more detail, including a study of course evaluations at Chalmers for the complete study year 2018/19. After discussing these findings, suggestions are made for additional/alternative evaluation methods, with a focus on one research-based method: the teaching practices inventory (TPI).

Problems with student course evaluations

It is common practice at Chalmers and many other universities to collect opinions on various aspects of course quality through student questionnaires. This practice is both valuable and problematic. Wankat and Oreovicz (2015) present a comprehensive list of various commonly criticized aspects and review relevant research connected with many of these aspects of student course/teacher evaluations.

It is worth noting that student satisfaction, course quality, teacher quality, and quality of teaching are separate but interrelated entities. Student satisfaction is what is easily measured, and is what is used as a substitute for more accurate but less used course quality measures.

Students' opinions on quality of teaching are collected in course evaluation questionnaires, but students are not trained in pedagogy, and therefore can only provide one-sided pictures of the learning experience (Deslauriers et al., 2019). In the best scenarios, teachers would use the best possible (research-based) teaching methods, informing their students of why they have chosen these methods, so that the students gain some of the insight needed to judge the choice and implementation of these teaching methods.

Course quality includes choice of subject content and level, as well as the design of constructively aligned goals (intended learning outcomes), teaching and learning activities, and examination methods. The students can definitely have opinions about all these aspects of course quality but they are generally not well qualified to evaluate these course design aspects.

Course quality is intertwined with quality of curriculum design at the program level, including prior knowledge and progression throughout the program. Students with and without the prior knowledge expected at the onset of a course are often treated as a homogeneous group when it comes to teaching and learning activities prescribed. Furthermore, the student opinions of course quality are not separated according to prior knowledge level, although student opinions of their own relevant prior knowledge are collected in the Chalmers student evaluation questionnaires. The anonymous nature of these questionnaires pre-empts the possibility of connecting student opinions to objective measures of individual prior knowledge.

The status of a specific course as obligatory or elective, as well as the level of the course (1st or 2nd cycle) also affects the students' overall impression of the quality of a course (Johnson et al., 2013). Other confounding factors include the type of course (mainly lecture-based as opposed to mainly laboratory or project-based). Malmqvist et al. (2018) have studied this at Chalmers.

Many universities rely heavily on student evaluations of teaching for evaluation of teaching staff (Linse, 2017). However, quality of teaching can depend heavily on the teaching context, and evaluation of teaching staff is meant to be based on the quality of an individual's ability, knowledge and performance, which can develop over time. This duality when it comes to personal attributes contra actions, reflects the difference between a fixed contra growth mindset (Dweck, 2015) when it comes to evaluation of teaching faculty.

Also somewhat problematic is the fact that courses are seldom taught by one teacher alone, although the examiner has the formal responsibility for leading the teaching team and seeing to it that good teaching practices are used by all within the course.

The Prioritized Operational Development document, (PVU, 2019) for Chalmers states the goal that at least 80% of courses at Chalmers should have a student satisfaction rating of 3.5 or more. When ratings fall below 3.0 on specific courses, an action plan is required. This is, at the moment, one of the few broadly monitored indicators of course quality used to guide Chalmers course development, even though it may seem to be a problematic indicator. (The other main indicator involves throughput.)

At Chalmers, there is an agreement with the employee organisation SACO which forbids the use of student "overall impression" data in the setting of salaries for teaching staff (SACO, 2019). This agreement rests upon insight into the differences between student satisfaction, course quality and teacher performance.

A recent statement from sociologists and other professional groups (ASA, 2019; Flaherty, 2019) pinpoints a number of reasons to be careful when interpreting the results of student feedback. One of these caution warnings relates to gender bias (Flegl, 2019), which motivated checking whether the gender of the examiner correlates with student ratings at Chalmers. Centra and Gaubatz (2000) have studied differences in how male and female students rate male or female instructors, over a number of

different disciplines with varying gender imbalance. Papadopoulou et al. (2019) found that student gender affected satisfaction with courses at Chalmers.

Results are presented below of a coarse look at the data on student satisfaction with courses held all across Chalmers during the academic year 2018/19 as it relates to the examiner's gender.

Gender bias test method

Student course evaluation data is available at Chalmers in a database showing the numerical average of students' answers to a number of set questions for each course. The students answer on a Likert scale from 1 to 5 where 5 corresponds to the most positive response (in most cases). The results of the question on overall impression (satisfaction) with the course are used as a numerical indication of course quality, with a specific goal of 3.5 or higher stipulated in steering documents at Chalmers (PVU, 2019).

In this study, the student overall satisfaction rating is analysed for all Chalmers undergraduate courses from the 2018/19 academic year with at least 10 respondents to the questionnaire and at least 50% response rate. These limits were chosen ad hoc to remove data with relatively high uncertainty. This data was accessed on 2019-09-18 through the QlikView system available to all employees at Chalmers.

The examiner of the course was identified through each course's course syllabus for the 2018/19 academic year published in the Study Portal system (open to the public). The gender of the examiner was identified using a subjective classification based on personal acquaintance, common naming conventions, or photographs published on the publicly available Chalmers employee presentation webpages. All course examiners were thus assigned to either the male or the female group, with no other attempt made to check the correctness of this subjective binary classification. Neither was any attempt made in this study to include any other possible gender identifications.

The data analysis is performed in two different ways. First, the mean rating for all courses with female examiners together was compared to the mean for all courses with male examiners, with a statistical test of the hypothesis that their means were the same within a 95% confidence (null hypothesis.)

The second analysis of the data involved checking whether the distribution of mean course ratings was similar for courses with female and male examiners. This analysis involved sorting the data into rating ranges of width 0.5 on the scale from 1.0 to 5.0 (the full range of the Likert scale). The proportion of courses with female or male examiners was then compared to the expected proportion in each rating range. The expected proportion is based on the gender distribution of the teaching faculty at Chalmers, 22% women and 78% men (Chalmers i siffror, 2018).

Gender bias test results

193 courses remained of the original 1098 courses after removal of courses with fewer than 10 respondents and/or lower than 50% response rate. Only 39 of the 193 courses included in this study had female examiners.

The mean rating for all 39 courses given by female examiners was 3.85 with a standard deviation of 0.65 while the mean for all 154 courses with male examiners was 3.93 with a standard deviation of 0.62. This difference is not statistically significant at the 95% confidence level (Råde and Westergren, 1995).

The number of courses in each rating range is tabulated separately for male and female examiners in Table 1. which also shows the proportion of female examiners in each group. Figure 1. provides a visual aid for comparison of the two distributions. The data suggests that Chalmers students favour courses given by male examiners when it comes to “top notch” ratings (4.5 and above). No significant gender difference can be seen in rating ranges below 4.5.

Table 1. Number of courses categorized by range of average student overall satisfaction rating and gender of examiner.

rating range	total nr. of courses	nr. of courses with male examiner	nr. of courses with female examiner	proportion of courses with female examiner (expected value 22%)
4.50 – 5.00	32	29	3	9%
4.00 – 4.49	67	50	17	25%
3.50 – 3.99	53	43	10	19%
3.00 – 3.49	20	15	5	25%
2.50 – 2.99	16	13	3	19%
2.00 – 2.49	4	4	0	(0%)
1.50 – 1.99	1	0	1	(100%)
1.00 – 1.49	0	0	0	
1.00 – 5.00	193	154	39	20%

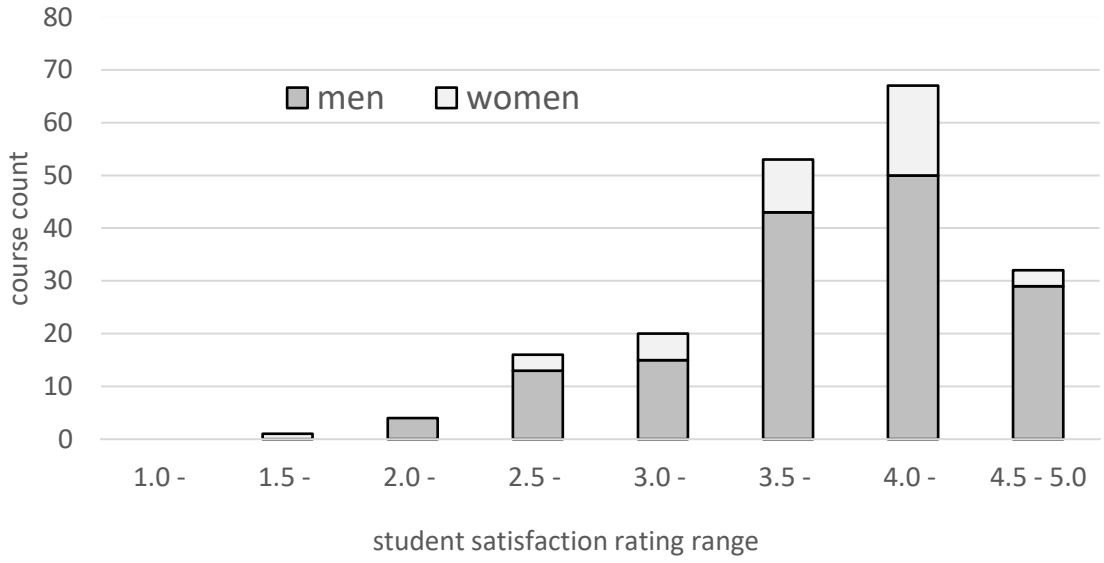


Figure 1. Number of courses with student satisfaction in each range of ratings of the “overall impression” question from the course evaluation questionnaires. Parameter: gender of examiner.

Discussion of gender bias

The data shown in Table 1. and Figure 1. indicate that there may be some trends in student overall satisfaction rating relating to the gender of the course examiner. However, the data is limited and further studies of more courses, with steps taken to increase the response rate on these student questionnaires might show other tendencies. Nevertheless, one can speculate as to whether the finding that the highest rating categories had the lowest proportion of female examiners might be a reflection of trends to rate excellence differently according to gender (van den Brink and Benschop, 2011).

Centra and Gaubatz (2000) found gender differences in teaching styles while studying gender bias in student evaluations of teaching. Different teaching styles are known to affect student learning, (Rust, 2013; Freeman *et al.* 2014) so might there be a gender difference in teaching quality? To address this question one would need a method to measure teaching styles, which could then be evaluated according to research findings on which teaching styles best support student learning in the particular area of study.

Recommendations for alternatives to evaluate course/teacher quality

Teaching practices that really work

The importance of student activating teaching practices has been highlighted by, among others, the 2001 Nobel laureate in physics, Carl Wieman, who advocates the application of a "scientific approach" to teaching (Wieman, 2007). Freeman et al. (2014) quantify the gain in student learning associated with active classroom techniques compared to the passive lecture setting.

The fact that many presently employed professors were taught by professors who were unknowing of the learning gains missed by passive lecturing styles is against us. The "old guard" thrived on passive lectures – that's why they are still around – and they tend to focus on developing research in their particular fields rather than taking in relevant research on education. This frustration has been reiterated time and again by Wieman himself and other leading advocates for change in university teaching, (Scott, 2017).

Teaching Practices Inventory

Based on research evidence of this type, Carl Wieman, together with his physicist wife, Sarah Gilbert, have developed a teaching quality evaluation tool, the Teaching Practices Inventory, for use in lecture-based mathematics and science courses at the university level (Wieman and Gilbert, 2014; Wieman, 2015). This is a questionnaire to the teacher responsible for the course, in which the degree of research-based teaching practices used in the course is evaluated. This Teaching Practices Inventory has been carefully validated, and shown to give results that correlate well with more direct measures of improvements in student learning.

The online questionnaire typically takes about 10 minutes to complete, and returns a numerical score which can be used to monitor the adoption of research-based teaching practices in a particular course. This relatively modest time investment should prove worthwhile as an input to the regular course evaluation meetings held after the end of each course at Chalmers. For those who might want to test the Teaching Practices Inventory, it is available online, including an automatic scoring of the answers provided (CWSEI).

Actual learning vs. feelings of learning

Examples of research-based teaching practices which contribute to better learning (and higher scores in the Teaching Practices Inventory) include different forms of actively engaging students in the classroom. One recent well-designed study of active learning in a physics lecture setting, using the same teachers and student groups as control groups, reported a negative correlation between actual learning (as measured with a test) and feelings of learning experienced by the students (Deslauriers et al., 2019).

Recommendations for Chalmers

At least for mainly lecture-based courses, it would be a positive step towards a more evidence-based quality measure to have the Teaching Practices Inventory score recorded alongside the numerical student evaluation results and examination rates in the database provided to support educational development at Chalmers, (presently in the system "QlikView"). The (hopefully positive) changes in the Teaching Practices Inventory scores over time would provide good input for individual teachers' pedagogical portfolios. Performance appraisal discussions would profit from a focus on more accurate quality measures, and it would be feasible to test the correlation between the Teaching Practice Inventory score and the overall student satisfaction scores that have come to be used at Chalmers in lieu of an evidence-based teaching quality measure.

Adaptation of the Teaching Practices Inventory

There is work still to be done to adapt the Teaching Practices Inventory to the specific educational structure at Chalmers, regarding e.g. study periods instead of semester-long courses, and adapting terminology for increasing the chance of homogeneous modes of interpretation of specific questions. It would also be fruitful to check the connection between the Teaching Practices Inventory questions and the present set of student survey questions, and possibly suggest changes or additions to the student survey questions to better reflect present knowledge of research-based teaching practices.

An alternative set of research-based teaching practice questions needs to be developed for non-lecture-based courses, such as project courses, lab courses or work experience-based courses. We would also need a clear marking of courses in our databases as to which courses can be classified as mainly lecture-based, project-based, lab-based or work-experience based, similar to the present marking of courses which teach sustainable development.

Further research questions

With a (fairly) robust system for measuring course quality in place, it would be interesting in the future to investigate whether this correlates with those examiners who hold a Diploma of Teaching and Learning in Higher Education.

It would also be an interesting exercise to check whether there are gender differences in the degree of research-based teaching practices used in courses at Chalmers. To this means, a timely introduction of such a new routine could provide data that might be available in time for the next KUL conference.

Challenge

It remains to be seen whether we (Chalmers) really mean that we intend to stay on course towards a world-class education. Why not begin by using scientifically sound methods to teach and evaluate teaching!

References

ASA (2019). Statement on Student Evaluations of Teaching. *American Sociological Association, September 2019*, 1-4.

van den Brink, M. & Benschop, Y. (2011). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization 19(4)* 507-524.

Centra, J. A. & Gaubatz, N. B. (2000). Is There Gender Bias in Student Evaluations of Teaching? *The Journal of Higher Education, 71(1)* 17-33.

Chalmers i siffror (2018). Årsberättelse Hållbarhetsrapport Årsredovisning Chalmers 2018. https://www.chalmers.se/sv/om-chalmers/arsberattelse-och-statistik/Sidor/default.aspx

CWSEI. Carl Weiman Science Education Initiative at the University of British Columbia. http://www.cwsei.ubc.ca/index.html

CWSEI Teaching Practices Inventory.
http://www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm
Available as an online survey, accessed 2019-10-07, at:
https://stanforduniversity.qualtrics.com/jfe/form/SV_9Slsclx3xWPASJ7?Q_JFE=qdg

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proc. Natl. Acad. Sci. U.S.A., 116* (39) 19251-19257.

Dweck, C. (2015). Carol Dweck Revisits the "Growth Mindset". *Education Week, Sep. 2015,* 1-4.

Flaherty, C. (2019). Sociologists and more than a dozen other professional groups speak out against student evaluations of teaching. *Inside Higher Ed, September 2019*, 1-4.

Flegl, M., Rosas, L.A.A. (2019). Do professor's age and gender matter or do students give higher value to professors' experience? Quality Assurance in Education, 27 (4), pp.511-532.

Freeman S, Eddy SL, McDonough M, Smith MK, Wenderoth MP, Okoroafor N, Jordt H (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA 111*, 8410–8415.

Johnson, M., Narayanan, A., & Sawaya, W. (2013). Effects of Course and Instructor Characteristics on Student Evaluation of Teaching across a College of Engineering. *Journal of Engineering Education,* 102 (2) pp. 289-318.

Linse, A.R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, pp.94-106.

Malmqvist, J., Almefelt, L., Bingerud, M. & Enelund, M. (2018). Lessons Learned from Student Satisfaction Surveys of CDIO Project Courses. Proceedings of the 14th International CDIO Conference, Kanazawa Institute of Technology, Kanazawa, Japan, June 28 – July 2, 2018.

Papadopoulou, P., Bingerud, M., Hulthén, E. & Enelund, M. (2019). Gender Differences in Student Satisfaction Surveys. *Proceedings of the 15th International CDIO Conference,* Aarhus University, Aarhus, Denmark, June 25 – 27, 2019.

PVU (2019). Prioritized Operational Development, steering document at Chalmers University of Technology. *Prioriterad verksamhetsutveckling 2020-2022*, dnr C 2019-0649. http://document.chalmers.se/download?docid=7a22346c-057a-4ce5-9737-788dada89492&lang=en

Rust, C. (2013). What we know. *Oxford Centre for Staff and Learning Development. Oxford Brookes University.* e-book available at: https://shop.brookes.ac.uk/product-catalogue/oxford-centre-for-staff-learning-development/books-publications/ebooks/what-we-know-by-chris-rust-ebook

Råde, L. & Westergren, B. (1995). *Mathematics Handbook for Science and Engineering.* Studentlitteratur, p.476

SACO (2019) Lönekriterier C2019-1799, D2.4.2. Underbilaga 1.1 till förhandlingsprotokoll daterat 2019-10-31. Available in Swedish only. Citation from page 5: *"Utfall av kursvärderingar får inte användas vid lönesättning eftersom det krävs en nyanserad och omfattande analys för att ge en rättvis bild av kvalitén på undervisningen och dess genomförande."* https://document.chalmers.se/download?docid=366f1acb-2041-479c-9c8a-4a46a0a2732f

Scott, S. (2017). Should We Lose the Lecture? An atomic physicist makes the case for active learning. *Stanford Magazine.* March/April 2017. https://stanfordmag.org/contents/should-we-lose-the-lecture

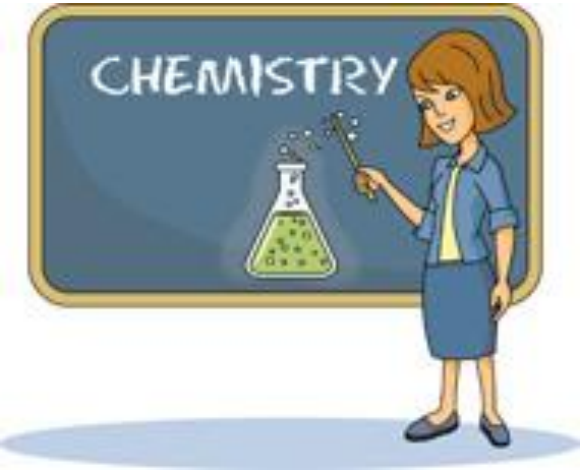
Wankat, P. & Oreovicz, F. (2015). *Teaching Engineering.* 2nd Ed., Purdue University Press, chapter 16.4

Wieman, C. (2007). Why not try a scientific approach to science education? Change: The Magazine of Higher Learning, 39:5, 9-15, DOI: 10.3200/CHNG.39.5.9-15

Wieman, C. & Gilbert, S. (2014). The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science. *Cell Biology Education – Life Sciences Education, 13(Fall),* 552-569.

Wieman, C. (2015). A Better Way to Evaluate Undergraduate Teaching. Carl Wieman (2015). *Change: The Magazine of Higher Learning,* 47:1, 6-15, DOI: 10.1080/00091383.2015.996077

A coarse look at course evaluations - coupling to course quality and gender bias



Sheila Galt, EER, CLS, Chalmers

KUL Conference 2020-01-10

Overview

- Quality of course/teaching/teacher
- Problems with student course evaluations

- Gender bias in Chalmers course evaluations?

- Alternatives to evaluate course/teacher quality
- Teaching practice inventory

- Moving forward at Chalmers

Quality of course/teaching/teacher – why evaluate?

- Course development
  - Learning goals
  - Learning activities
  - Learning evaluation

} All constructively alligned!

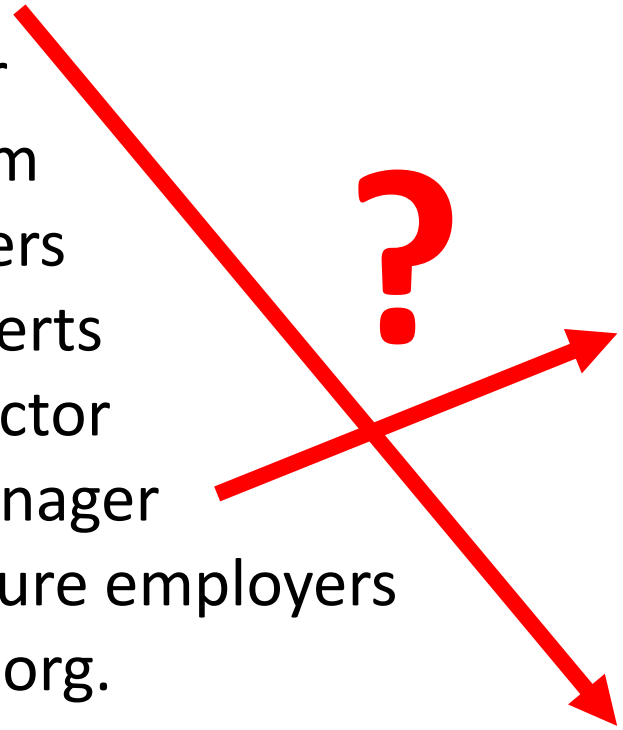
- Program development
  - Curriculum design
  - Prerequisite knowledge
  - Progression through learning sequences
  - Professional skills development
- Teacher development
  - Teaching practices
  - Attitudes and priorities

Who or what are we evaluating?

- Different but interrelated:
  - Quality of course
  - Quality of learning
  - Quality of teaching
  - Quality of teacher
    - Main teacher (examiner)
    - Members of teaching team (other lecturers, TAs, lab assistants)

Who can/should evaluate what aspects?

- Students
- Main teacher
- Teaching team
- Teacher's peers
- Teaching experts
- Program director
- Teacher's manager
- Students' future employers
- Government org.
- Academic peers



- Complete course experience
- Course design, lesson content, exam results
- Practicalities
- Pedagogical content knowledge
- Teaching practices
- Learning goals and curriculum
- Time, budget, professional development
- Employee performance
- Rules, regulations, laws
- Validity of examination

When to evaluate course quality?

Before or after:

- writing final exam?
- course grades delivered?

Student course evaluation questionnaire:

- opens 1st day of next study period
- closes after 2 weeks (SP4 – over summer)

exam results usually delivered during or after questionnaire open time span

- Directly after end of course – best time for teacher to reflect
- Before course plan revision deadline (jan/feb) – huge time lag SP3 and SP4
- Review close to start of next year's course

- When professionally applying knowledge, skills and attitudes in diverse future employment scenarios

Problems with student course evaluations

- Students well qualified to evaluate some but not all quality aspects
- Statistically problematic for small courses and/or low response rates
- Ease of data collection obscures alternative evaluation methods
- Student satisfaction used as (sole) basis for certain decisions
  - Requirement of written action plan (when general impression < 3.0)
  - Department level quality follow-up … (but not basis for quality-based funding)
  - Should not affect salary of teachers (SACO agreement with Chalmers)
- Psychosocial work environment for teachers affected by student ratings and comments (even after censuring)
- Risk of bias connected to student/teacher gender (or other characteristics)

Factors found to affect student course ratings

- Obligatory or elective course
- Level of course (Bachelor or Master level)
- Type of course (mainly lectures, project-based or mainly labs)
- Ethnic background or native language of teacher
- Gender of student or teacher

Wankat, P. & Oreovicz, F. (2015). *Teaching Engineering.* 2nd Ed., Purdue University Press, chapter 16.4

Gender bias in student evaluations of teaching

- Evidence of same gender preferences
  - Male-dominated student body implies positive bias for male teachers

- Evidence of gender differences in teaching styles
  - Men lecture more often, women use more in-class discussion

Centra, J. A. & Gaubatz, N. B. (2000).
Is There Gender Bias in Student Evaluations of Teaching?
The Journal of Higher Education, 71(1) 17-33.

Papadopoulou, P., Bingerud, M., Hulthén, E. & Enelund, M. (2019).
Gender Differences in Student Satisfaction Surveys.
Proceedings of the 15th International CDIO Conference, Aarhus University, Aarhus, Denmark, June 25 – 27, 2019.

Gender bias in Chalmers course evaluations?

- Study of overall impression rating for 2018/19
- All Chalmers courses with ”enough” data in QlikView
  - At least 50% response rate and at least 10 respondents
  - 193 courses included in this study (out of 1098)

iskreta system	2018/2019	LP1	51	37,5%	4,16
amik och strömningsmekanik	2018/2019	LP3	3	37,5%	4,67
öka arkitektur 1	2018/2019	LP4	29	37,7%	3,10
elanalys	2018/2019	LP3	91	37,8%	3,99
k analys i en variabel	2018/2019	LP2	31	37,8%	3,71
gonomi	2018/2019	LP2	25	37,9%	4,20
timering	2018/2019	LP3	25	37,9%	3,60
material, fortsättningskurs	2018/2019	LP1	11	37,9%	3,91
nentmetoden - strukturer	2018/2019	LP3	33	37,9%	4,06
toder	2018/2019	LP3	19	38,0%	4,89
och organisation i byggsektorn	2018/2019	LP2	84	38,0%	4,00
matematisk analys	2018/2019	LP1	119	38,0%	3,95
nterat programmeringsprojekt	2018/2019	LP1	43	38,1%	3,56
ministration	2018/2019	LP3	16	38,1%	4,13
fysik och byggnadsakustik, civilingenjör	2018/2019	LP2	16	38,1%	2,50
nteraction	2018/2019	LP3	16	38,1%	3,63
erade material	2018/2019	LP4	8	38,1%	4,50

- Gender of examiner identified from Study Portal course plan (syllabus)
- (Gender of students not studied here)

- Two null hypotheses to test:
  for groups of courses with male or female examiner:
  - No difference in average of course ratings between groups
  - No difference in distributions of course ratings between groups

Results of 2018/19 study – average ratings

- 193 courses included in this study:

  39 of a total of 193 courses had a female examiner (20%)

  proportion of female teaching faculty at Chalmers: 22%

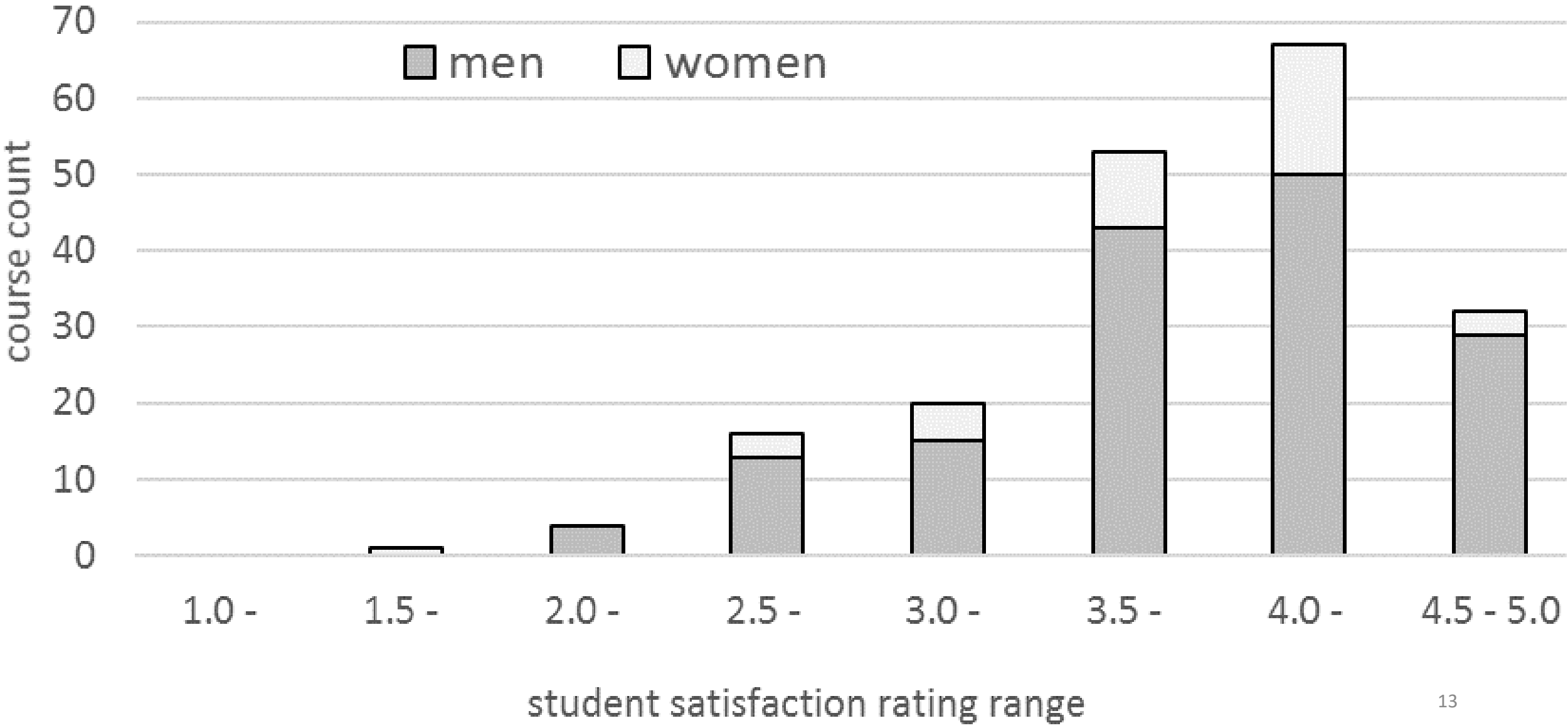
- No significant difference between average of course ratings

Female examiner		Male examiner	
Mean	Standard dev.	Mean	Standard dev.
3.85	0.65	3.93	0.62

Results of 2018/19 study – distribution of ratings

rating range	total nr. of courses	nr. of courses with male examiner	nr. of courses with female examiner	proportion of courses with female examiner
4.50 – 5.00	32	29	3	9%
4.00 – 4.49	67	50	17	25%
3.50 – 3.99	53	43	10	19%
3.00 – 3.49	20	15	5	25%
2.50 – 2.99	16	13	3	19%
2.00 – 2.49	4	4	0	(0%)
1.50 – 1.99	1	0	1	(100%)
1.00 – 1.49	0	0	0	
1.00 – 5.00	**193**	**154**	**39**	**20%**

Results of 2018/19 study – distribution of ratings



Observations, speculations and questions raised

- Distributions of course ratings are not significantly gender skewed

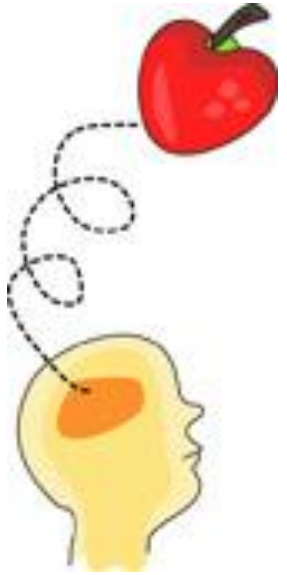
- Do male teachers at Chalmers get more ”top notch” ratings?
- Does gender bias in ”excellence” perceptions affect student ratings?

- Would analysis of previous study years give similar results?
- Are changes in course ratings connected to changes in teaching and learning?

- Do students give higher or lower ratings to courses taught with evidence-based effective teaching practices?
- Do female and male teachers differ in their use of evidence-based effective teaching practices?

Alternatives to evaluate course/teacher quality

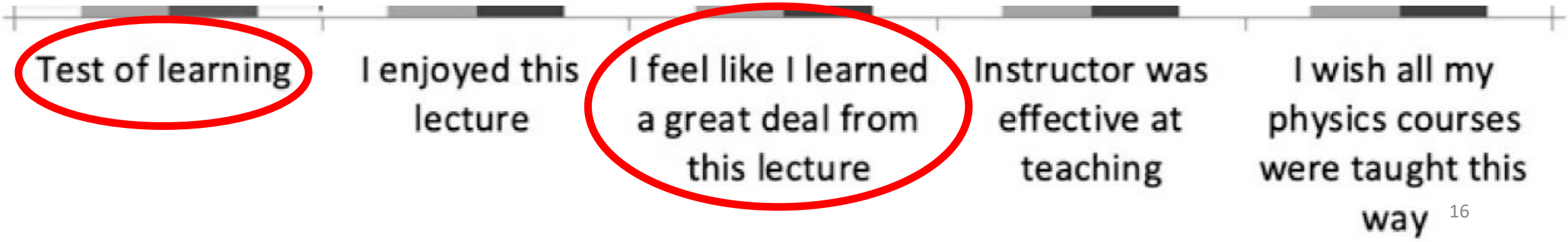
- Actual (not just intended) learning outcome analysis
  - Examination results
  - Test of prerequisite knowledge in next course in progression
  - Alumni reflection
  - Employer satisfaction
- Classroom observations (auskultations)
  - With pre- and post observation discussions
  - Observer may be peer or teaching expert
- Course analysis in relation to:
  - Course design, program design
  - Evidence-based effective teaching practices



Evidence-based effective teaching practices

- Active classroom techniques vs. passively listening to lectures
  - Evidence for better actual learning in active classrooms settings
- Freeman S, Eddy SL, McDonough M, Smith MK, Wenderoth MP, Okoroafor N, Jordt H (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA 111*, 8410–8415

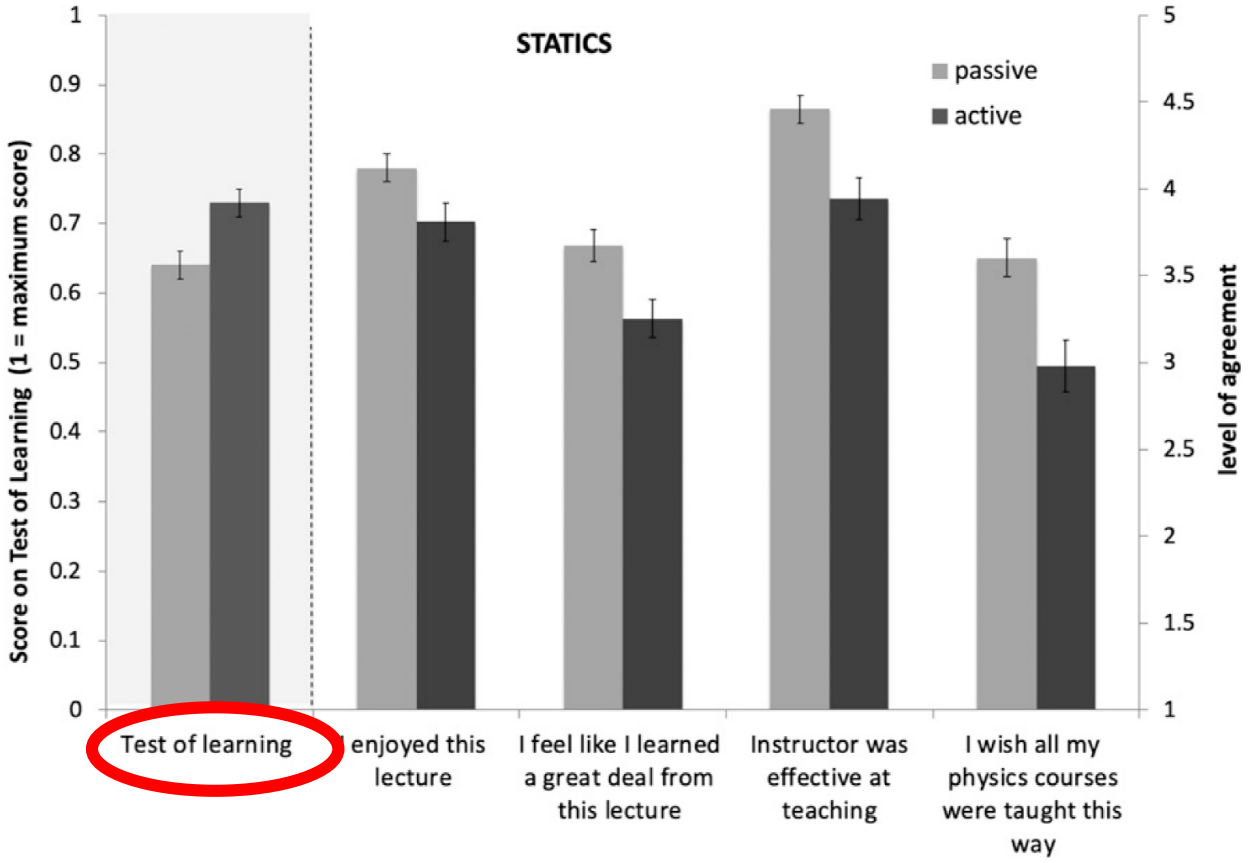
- Evidence for less "feelings of learning" among students in active settings
  - Note bar graph labels below, see two full bar graphs of results on next slide
- Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proc. Natl. Acad. Sci. U.S.A., 116* (39) 19251-19257.

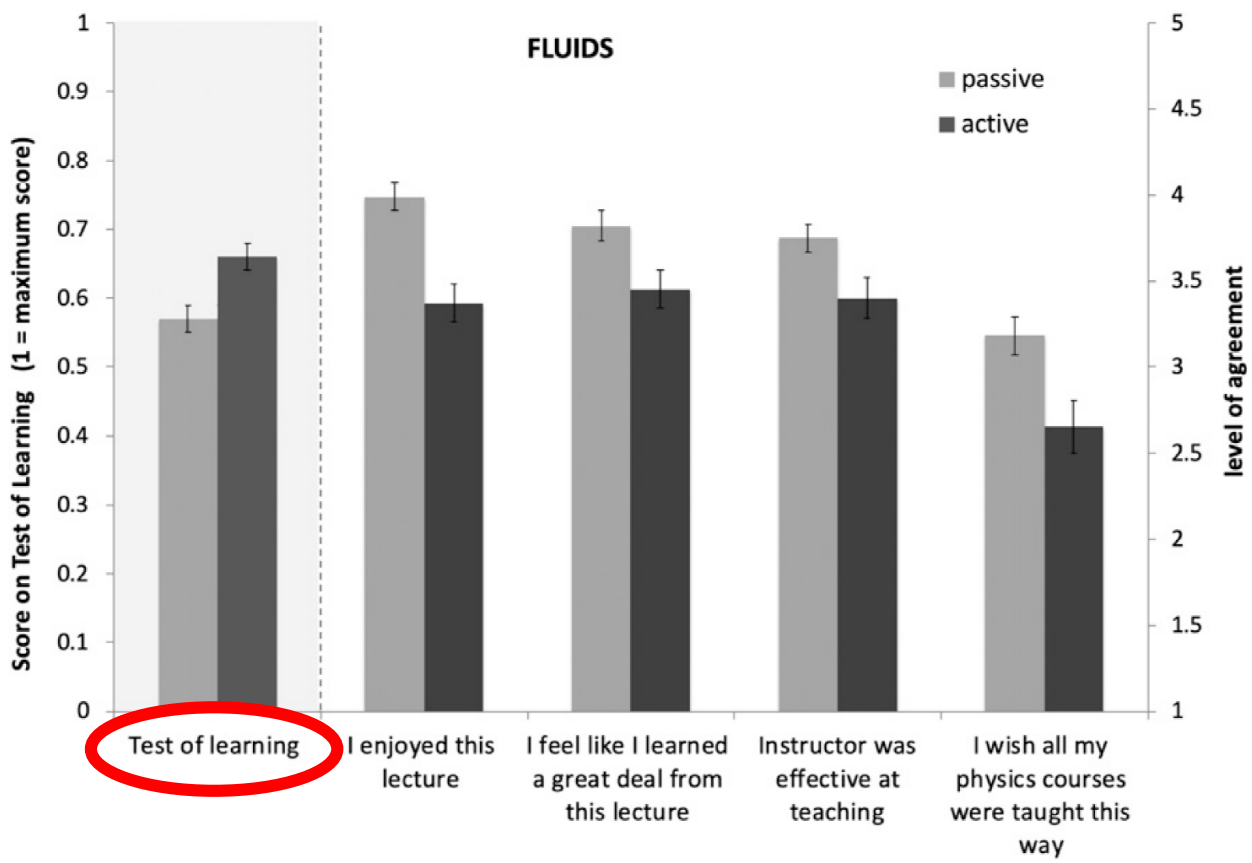


Actual learning vs feeling of learning – Deslauriers

Two different physics topics, two ”good” teachers, crossover study design

gray: traditional lecture (passive) black: active classroom





Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019).
Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom.
Proc. Natl. Acad. Sci. U.S.A., 116 (39) 19251-19257.

Teaching practices inventory (TPI)

- Example of a Teaching Practices Inventory
  developed at UBC (Vancouver, Canada) by
  Carl Wieman (Physics Nobel Laureate 2001) and Sarah Gilbert
  based on ”scientific approach” to teaching

Wieman, C. (2007). Why not try a scientific approach to science education? Change: The Magazine of Higher Learning, 39:5, 9-15, DOI: 10.3200/CHNG.39.5.9-15

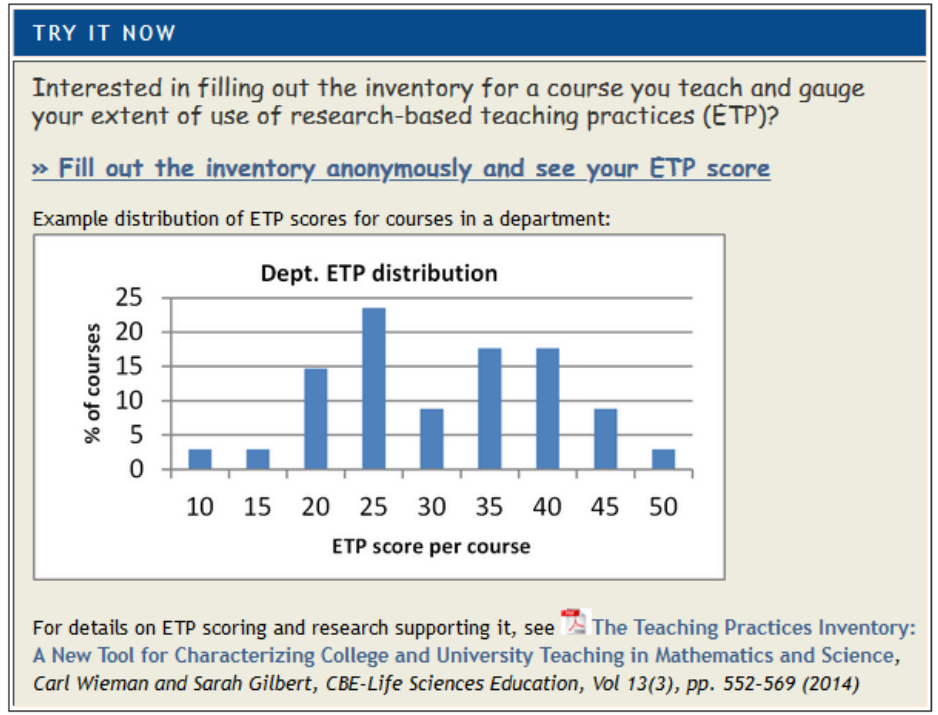
Wieman, C. & Gilbert, S. (2014). The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science. *Cell Biology Education – Life Sciences Education, 13(Fall),* 552-569.

How to use a teaching practices inventory (TPI)

- Self evaluation by teacher
- Standard questionnaire
- Suitable for mainly lecture-based courses
- Could be adapted for Chalmers use

- Available online (with automatic scoring) at:

CWSEI Carl Wieman Science Education Initiative
http://www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm



Examples of questions included in TPI:

- Students read/view material on upcoming class session and complete assignments or quizzes on it shortly before class or at beginning of class
- Average number of times per class: show demonstrations, simulations, or video where students first record predictions (write down, etc.) and then afterwards explicitly compare observations with predictions
- Reflective activity at end of class, e.g. "one-minute paper" or similar (students briefly answering questions, reflecting on lecture and/or their learning, etc.)
- Homework/problem sets assigned and contributed to course grade at intervals of 2 weeks or less
- There are Instructor-TA meetings every two weeks or more frequently where student learning and difficulties, and the teaching of upcoming material are discussed

Moving forward at Chalmers - suggestions

- Help all students, teaching colleagues and management understand the power of research based, effective teaching methods.
- Choose a suitable teaching practices inventory (possibly Wieman & Gilbert) and adapt it to Chalmers educational system.
- Introduce the TPI tool (gradually, pilot first) as a complement to student course evaluations and publish results in QlikView.
- Develop formal and informal TPI discussions
  - Department level follow-up
  - Employee appraisal discussions
  - Pedagogical development seminars
  - Reflections on TPI scores in pedagogical portfolios
  - Starting point for peer (or pedul) observation discussions

Further research questions raised

- Do TPI scores and student passing rates and/or grades correlate?
- Do TPI scores and student overall impression correlate?
- Is there a measurable gender difference in TPI scores?
- Do TPI scores correlate with formal teacher qualifications such as the Diploma och Higher Education?
- Is there evidence that the introduction of a TPI system can act as a driving force to hasten the adoption of research-based teaching practices?

Challenge

Do we mean what we say in striving for:

”world class education” at Chalmers

with scientifically sound teaching methods

and scientifically sound methods to evaluate teaching

?

Thank you for listening!

That’s my 2 cents worth!

A penny for your thoughts?